1Г

УДК 004.93

Е.А. Гофман, А.А. Олейник, С.А. Субботин
Запорожский национальный технический университет, Украина
olejnikaa@gmail.com

Эволюционный метод синтеза деревьев решений*

Исследован эволюционный подход к построению деревьев решений. Разработан метод редукции обучающей выборки, позволяющий исключать малозначимые и малоинформативные признаки. Предложен эволюционный метод построения деревьев решений. Проведены эксперименты по синтезу диагностических моделей на основе предложенных методов.

Введение

Решение задач технического и биомедицинского диагностирования связано с необходимостью построения моделей исследуемых объектов, процессов или систем [1].

В качестве диагностических моделей эффективно могут использоваться деревья решений [2-5], которые являются достаточно популярным направлением вычислительного интеллекта и строятся на основе индуктивного обучения по примерам. Деревья решений представляют собой графовые модели, узлы которых реализуют функции принятия решений на основе значений входных переменных, а листовые узлы содержат значения выходной переменной, соответствующие условиям внутренних узлов. Такие модели хорошо интерпретируются людьми-специалистами в прикладных областях, которые, как правило, не знакомы с методами и моделями искусственного интеллекта. Деревья решений позволяют эффективно решать задачи классификации и прогнозирования, обеспечивая при этом высокую точность.

Для синтеза деревьев решающих правил в настоящее время используют методы CART, C4.5, ID3 и др. [2]. Однако такие методы используют жадную стратегию поиска. Если один раз признак был выбран, и по нему выполнено разбиение на подмножества, то метод не может вернуться назад и выбрать другой атрибут, который привел бы к лучшему разбиению, вследствие чего в результате часто синтезируются деревья решений, не обеспечивающие приемлемый уровень аппроксимации [2-5].

Поэтому построение деревьев решений целесообразно выполнять на основе эволюционного поиска [1], [6-8], представляющего собой набор стохастических методов оптимизации, не использующих жадную стратегию поиска, не требующих вычислений значений производных оптимизируемого функционала и обладающих механизмами выхода из локальных оптимумов.

Целью статьи является разработка эволюционного метода синтеза деревьев решений, позволяющего выполнять построение древовидной диагностической модели без использования жадной стратегии поиска.

---

*Работа выполнена в рамках госбюджетной НИР «Информационные технологии автоматизации распознавания образов и принятия решений для диагностики в условиях неопределенности на основе гибридных нечетkологических, нейросетевых и мультиагентных методов вычислительного интеллекта» кафедры программных средств Запорожского национального технического университета (номер государственной регистрации 0109U007673).

Постановка задачи

Пусть задана выборка данных $< X, Y >$, где $X = \{X_i\}$ – набор значений признаков, характеризующих рассматриваемый объект или процесс; $Y = \{y_p\}$ – массив значений выходного параметра в заданной выборке; $X_i = \{x_{ip}\}$ – i-й признак в выборке, $i = 1, 2, ..., L$; x_{ip} – значение i-го признака для p-го экземпляра выборки, $p = 1, 2, ..., m$; y_p – значение прогнозируемого параметра для p-го экземпляра; L – общее количество признаков в исходном наборе; m – количество экземпляров выборки.

Необходимо аппроксимировать зависимость отклика Y от набора факторов X в виде модели, представленной деревом решений $T = \{t_k\}$ таким образом, чтобы значение ошибки прогнозирования или классификации E построенной модели было минимальным, где $t_k = < c_k, l_k, r_k >$ – k-й узел дерева T, представляющий собой структуру, в которой c_k – функция принятия решений на основе значений входных переменных (в случае, если узел является внутренним) или значение выходной переменной (для внешних узлов), l_k и r_k – ссылки на левого и правого потомков k-го узла, соответственно, представляющих собой структуры, аналогичные t_k.

Тогда задача построения дерева решений $T = \{t_k\}$ по заданной выборке $< X, Y >$ заключается в идентификации узлов $t_k = < c_k, l_k, r_k >$ так, чтобы $E \rightarrow \min$.

Редукция обучающей выборки

Известно [1], [2], [6], что обучающая выборка может содержать избыточные или малозначимые признаки, которые не только усложняют синтезируемую модель, но и затрудняют процесс ее построения. Кроме того, в обучающем множестве могут находиться избыточные экземпляры, исключение которых позволит сократить размер выборки, и, следовательно, повысить обобщающие свойства построенной на ее основе модели.

Поэтому до синтеза диагностических моделей целесообразно выполнять редукцию обучающей выборки, исключая избыточные признаки и экземпляры. Поэтому разработанный метод редукции обучающей выборки предлагается выполнять в два этапа:

– редукция признаков;

– редукция экземпляров.

Для редукции признаков предлагается использовать эволюционный поиск, основанный на оптимизации случайного множества решений с различными оценками, а не одного решения, что позволяет синтезировать множество новых решений на основе старых субоптимальных решений. Однако методы эволюционного поиска являются высокоитеративными, а время их работы сильно зависит от начальных условий поиска. Поэтому эффективность работы традиционных методов эволюционного поиска является недостаточно высокой.

Для повышения эффективности методов эволюционного поиска при решении задачи отбора информативных признаков предлагается анализировать информативность признаков на этапе инициализации параметров эволюционного метода путём применения известных критериев оценивания индивидуальной информативности признаков (коэффициент парной корреляции, коэффициенты корреляции знаков, энтропия признаков, теоретико-информационный критерий и др.).

При инициализации начального множества решений предлагается увеличить вероятность включения тех признаков в хромосому, чьи оценки индивидуальной значимости выше средней.

При создании новых решений с помощью оператора равномерного скрещивания априорная информация о значимости признаков может быть использована следующим образом. Вначале определить пороговое значение информативности признаков. Затем создать маску скрещивания, установив «1» для тех признаков, оценки информативности которых выше порога, и «0» – для признаков, оценки информативности которых равны или ниже порога. С целью увеличения разнообразия популяции после формирования маски скрещивания можно случайным образом изменить в ней некоторые разряды.

В операторе мутации априорные сведения об уровне информативности признаков можно использовать посредством усиления вероятности мутации признаков, имеющих низкие оценки индивидуальной информативности и ослабления вероятности мутации признаков с высокими оценками индивидуальной информативности. Это позволит закрепить признаки, наиболее сильно влияющие на целевой параметр модели, и сконцентрировать поиск на переборе сочетаний признаков с меньшей информативностью.

На этапе редукции экземпляров из обучающей выборки исключаются экземпляры, не существенные для построения диагностической модели. Очевидно, что наибольшую значимость для синтеза моделей оказывают экземпляры, расположенные в центрах и на границах классов или кластеров. Поэтому для оценивания значимости экземпляров обучающей выборки предлагается использовать следующие критерии:

1) критерий индивидуальной информативности p-го экземпляра относительно внутренней границы, который будет принимать значения в диапазоне от нуля до единицы, а его значение будет тем больше, чем ближе экземпляр расположен к границе между разными классами:

$$IC_{1p} = \frac{1}{m-1}\sum_{s=1}^{m}\left\{ e^{-\alpha(y_p, y_s)\sum_{i=1}^{L}(x_{ip}-x_{is})^2} \,\middle|\, p \neq s, y_p \neq y_s \right\},$$

где $\alpha(k,q) = \dfrac{1}{\dfrac{1}{m_k m_q}\sum\limits_{p=1}^{m}\sum\limits_{s=1}^{m}\left\{\sum\limits_{i=1}^{L}(x_{ip}-x_{is})^2 \,\middle|\, (y_p = k, y_s = q) \vee (y_p = q, y_s = k)\right\}}$;

2) критерий индивидуальной информативности p-го экземпляра относительно его удаленности:

– относительно границы класса:

$$IC_{2p} = 1 - \min_{s=1,2,\ldots,m}\left\{ e^{-\beta_{y_p}\sum_{i=1}^{L}(x_{ip}-x_{is})^2} \,\middle|\, p \neq s, y_p = y_s \right\},$$

где $\beta_k = \dfrac{1}{\max\limits_{\substack{p=1,2,\ldots,m;\\ s=p+1,\ldots,m}}\left\{\sum\limits_{i=1}^{L}(x_{ip}-x_{is})^2 \,\middle|\, y_p = k, y_s = k\right\}}$;

– относительно внешней границы выборки:

$$IC_{3p} = 1 - \min_{s=1,2,\ldots,m}\left\{ e^{-\gamma\sum_{i=1}^{L}(x_{ip}-x_{is})^2} \,\middle|\, p \neq s \right\},$$

где $\gamma = \dfrac{1}{\max\limits_{\substack{p=1,2,\ldots,m;\\ s=p+1,\ldots,m}}\left\{\sum\limits_{i=1}^{L}(x_{ip}-x_{is})^2\right\}}$.

Значение данного критерия будет тем больше, чем дальше экземпляр расположен по отношению к остальным экземплярам класса или выборки, т.е. критерий будет выявлять редкие нетипичные случаи, а также экземпляры, расположенные на внешних границах классов;

3) критерий индивидуальной информативности p-го экземпляра относительно его близости к центру класса (кластера) определим как:

$$IC_{4p} = \frac{1}{m_{y_p} - 1} \sum_{s=1}^{m} \left\{ \mathrm{e}^{-\sum_{i=1}^{L}(x_{ip}-x_{is})^2} \middle| p \neq s, y_p = y_s \right\}.$$

Данный критерий будет принимать значения в диапазоне от нуля до единицы: его значение будет тем больше, чем ближе экземпляр расположен к центру класса (кластера).

Для редукции экземпляров определим интегральный показатель информативности p-го экземпляра определим как:

$$IC_p = \max\{IC_{1p}, IC_{2p}, IC_{3p}, IC_{4p}\}.$$

Данный критерии будет принимать значения в диапазоне от нуля до единицы: чем больше будет его значение, тем значимее p-й экземпляр для построения модели, поскольку он либо находится на границе между классами, либо является уникальным наблюдением, либо находится на внешней границе класса, либо близок к центру кластера.

Предложенный метод редукции обучающей выборки, в отличие от существующих, не только исключает малоинформативные признаки, но и удаляет из выборки экземпляры, не существенные для построения диагностических моделей, что позволяет синтезировать более простые модели, обеспечивающие приемлемые аппроксимационные и обобщающие свойства, затрачивая при этом меньше времени на процесс их идентификации.

Эволюционный метод построения деревьев решений

Как отмечено выше, известные методы построения деревьев решений используют жадную стратегию поиска, не позволяющую в процессе поиска заменять атрибуты, по которым уже выполнено разбиение, на другие, более эффективные. Это в некоторых случаях приводит к тому, что синтезированные деревья решений не обладают приемлемыми обобщающими способностями.

Поэтому для синтеза деревьев решений предлагается использовать эволюционный поиск, который на основе стохастического подхода позволяет перебирать различные комбинации, не зацикливаясь на исследовании субоптимальных областей.

В разработанном эволюционном методе построения деревьев решений предлагается информацию о деревьях решений хранить в виде хромосом, представленных с помощью древовидной структуры данных. Для оценивания хромосом используется функция, отражающая информацию как о качестве аппроксимации с помощью текущего дерева решений, так и о его сложности, что позволяет строить деревья решений с приемлемыми обобщающими и аппроксимционными свойствами, содержащие при этом незначительное количество узлов, что в свою очередь повышает интерпретабельность синтезированных моделей. Для создания нового множества решений с помощью эволюционных операторов скрещивания и мутации предлагается использовать операторы, аналогичные операторам генетического программирования, поскольку такие операторы эффективно генерируют новые решения, представленные в виде древовидных структур.

Предложенный эволюционный метод синтеза деревьев решений состоит из следующих этапов.

Этап 1. Инициализация. На данном этапе необходимо редуцировать обучающую выборку $< X, Y >$, выделив наиболее существенные признаки и экземпляры.

После этого создается начальное множество решений, представленных в виде хромосом H_j, $j = 1, 2, \ldots, N$, где N – количество хромосом. При этом каждая хромосома H_j соответствует дереву решений T_j.

Для кодирования хромосомы (представления дерева решения в виде хромосомы) целесообразно использовать структуру данных в виде бинарного дерева $T = \{t_k\}$, в узлах $t_k = < c_k, l_k, r_k >$ которого расположены функции c_k принятия решений на основе значений входных переменных (в случае, если узел является внутренним) или значение выходной переменной (для внешних узлов), а также ссылки l_k и r_k на левого и правого потомков, соответственно, представляющих собой структуры, аналогичные t_k.

Этап 2. Оценивание текущего набора решений. В качестве целевой функции целесообразно использовать функцию, учитывающую:

– ошибку прогнозирования или классификации $E(T_j)$ с помощью дерева T_j, соответствующего оцениваемой хромосоме H_j;

– сложность синтезированного дерева $S(T_j)$.

Учитывая оба критерия, предлагается использовать такую целевую функцию:

$$f(H_j) = \alpha_1 E(T_j) + \alpha_2 S(T_j) \to \min,$$

где α_1 и α_2 – коэффициенты, учитывающие значимость каждого из целевых критериев ($E(T_j)$ и $S(T_j)$, соответственно), $\alpha_1 + \alpha_2 = 1$.

Для оценивания сложности дерева можно использовать следующую формулу:

$$S(T_j) = N(T_j) + r(T_j) + \eta(T_j),$$

где $N(T_j)$ – количество узлов в дереве T_j;

$r(T_j)$ – количество ребер в дереве T_j;

$\eta(T_j)$ – глубина дерева T_j.

Этап 3. Проверка критериев останова поиска. В качестве таковых критериев можно использовать следующие:

– достижение максимального количества итераций поиска $Iter_{\max}$;

– достижение приемлемого значения целевой функции $f^* \le \varepsilon$, где ε – заданное пользователем максимально допустимое значение целевого критерия; f^* – минимальное значение $f(H_j)$ целевой функции на текущей итерации.

Если хотя бы один из критериев останова достигнут, то завершить поиск. При этом считается, что оптимальное дерево решений представлено в виде хромосомы H^*, значение целевой функции которой равно f^*.

В случае если критерии останова не удовлетворены, выполнить переход к этапу 4.

Этап 4. Отбор. Выбор хромосом для скрещивания и мутации предлагается выполнять пропорционально значению их целевой функции. Для этого рассчитывается среднее значение целевой функции в текущей популяции по формуле:

$$f_{\text{ср}} = \frac{1}{N} \sum_{j=1}^{N} f_j .$$

После этого для каждой хромосомы H_j вычисляется величина $p_о(H_j)$ как отношение значения целевой функции $f(H_j)$ к среднему значению целевой функции в популяции $f_{ср}$:

$$p_о(H_j) = \frac{f(H_j)}{f_{ср}}.$$

Величина $p_о(H_j)$ определяет вероятность хромосомы H_j быть отобранной для скрещивания и мутации. При этом хромосома H_j копируется *k* раз в массив особей, допущенных к воспроизводству, где величина *k* вычисляется по формуле:

$$k = \begin{cases} \text{целое}\left(p_о\left(H_j\right)\right), \text{ если}\left(p_о\left(H_j\right) - \text{целое}\left(p_о\left(H_j\right)\right)\right) < rand(0;1), \\ \text{целое}\left(p_о\left(H_j\right)\right) + 1, \text{в противном случае;} \end{cases}$$

где $rand(0;1)$ – случайно сгенерированное число в диапазоне (0;1).

Этап 5. Скрещивание. Основной особенностью реализации генетических операторов скрещивания и мутации при построении деревьев решений в предложенном методе является поддержка синтаксической целостности получаемых решений (хромосом). Оператор скрещивания предлагается выполнять следующим образом. Вначале выбираются случайным образом на каждом из родительских деревьев одна или несколько точек (узлов дерева решений). Точки разрыва выбираются случайным образом для каждого родителя отдельно, что приводит к тому, вследствие выбора разных точек разрыва возможны ситуации порождения одинаковыми родителями разных потомков. После выбора точек скрещивания выполняется генерация потомков. Для этого необходимо обменять у родительских деревьев фрагменты согласно выбранным точкам разрыва.

Этап 6. Мутация. Для изменения деревьев решений, выбранных для мутации, необходимо случайным образом выбрать мутирующий узел дерева, после чего выполнить мутацию одним из следующих способов:

– мутация с увеличением – в выбранном узле построить новую ветвь дерева, увеличив таким образом дерево;

– мутация со сжатием – удалить выбранный узел и все его потомки, уменьшив тем самым размер дерева;

– мутация с заменой – после выбора двух узлов дерева, выполняется обмен соответствующих ветвей.

Этап 7. Формирование нового поколения. На данном этапе создается новое множество решений из хромосом, созданных в результате скрещивания и мутации, а также из наиболее приспособленных хромосом предыдущего поколения (хромосом с наилучшими оценками целевой функции). После этого выполняется переход к этапу 2.

Предложенный эволюционный метод синтеза деревьев решений основан на стохастическом подходе и не использует жадную стратегию поиска, позволяет строить деревья решений с приемлемыми обобщающими и аппроксимационными свойствами, содержащие при этом незначительное количество узлов, что в свою очередь повышает интерпретабельность синтезированных моделей.

Эксперименты и результаты

Для проверки эффективности использования предложенных методов редукции обучающей выборки и синтеза деревьев решений на основе эволюционного подхода была выполнена их программная реализация.

После чего решалась задача синтеза диагностической модели коэффициента упрочнения деталей авиадвигателей от различных параметров.

В настоящее время для упрочнения лопаток, валов и дисков газотурбинных двигателей широко применяется алмазное выглаживание [6], которое, в отличие от ручного полирования полотна, позволяет получать более благоприятные и стабильные свойства поверхностного слоя за счет пластической деформации и механизации процесса упрочнения, что повышает надежность и ресурс обрабатываемых деталей.

Показателем эффективности упрочнения алмазным выглаживанием деталей, работающих при циклических нагрузках, является коэффициент упрочнения β^{y}, который представляет собой один из важнейших факторов при расчете запаса прочности деталей газотурбинных авиадвигателей [6].

В настоящее время расчет коэффициента упрочнения β^{y} обычно производится по результатам испытаний на усталость стандартных образцов [6], что является дорогостоящей и не всегда практически осуществимой задачей. Поэтому актуальной является задача построения математической модели коэффициента упрочнения, позволяющей оценивать его значения без проведения испытаний деталей авиадвигателей на усталость.

Эффективность алмазного выглаживания в значительной мере зависит от выбранных режимов, физико-механических и геометрических характеристик упрочняемых деталей и деформирующего инструмента.

В качестве факторов, наиболее полно отражающих процесс алмазного выглаживания деталей, используют: x_1 – твердость материала (HB, МПа); x_2 – среднее контактное давление в зоне деформирования (q, МПа); x_3 – полуось эллипса касания в зоне упругого контакта (a, мм); x_4 – подача при выглаживании (s, мм/об.); x_5 – предел прочности материала (σ_B, МПа); x_6 – предел текучести материала ($\sigma_{0,2}$, МПа); x_7 – показатель деформационного упрочнения (n); x_8 – теоретический коэффициент концентрации напряжений от следов обработки ($\alpha_\sigma^{\text{техн}}$); x_9 – параметр исходной шероховатости детали (R_{a1}, мкм); x_{10} – сила выглаживания ($P_{\text{у}}$, Н); x_{11} – радиус сферы алмазного инструмента ($R_{\text{сф}}$, мм); x_{12} – параметр шероховатости после выглаживания детали (R_{a2}, мкм); x_{13} – теоретический коэффициент концентрации напряжений натурной детали или образца (α_σ); x_{14} – диаметр образца в опасном сечении (d, мм); x_{15} – радиус скругления галтели или надреза (r, мм); x_{16} – относительный градиент первого главного напряжения ($\overline{G}$, мм$^{-1}$). Таким образом, обучающая выборка содержала 16 признаков и 189 экземпляров.

На первом этапе вычислительных экспериментов с помощью предложенного метода редукции обучающей выборки были исключены малоинформативные признаки (x_3, x_6, x_{13}, x_{15}) и 52 малозначимых экземпляров.

Далее проводился синтез деревьев решений с помощью методов CART, C4.5 и предложенного эволюционного метода. Результаты, полученные с помощью различных методов, приведены в табл. 1.

Таблица 1 – Сравнение различных методов синтеза деревьев решений

№ п/п	Метод синтеза дерева решений	Характеристики синтезированного дерева			
		$E(T_j)$	$N(T_j)$	$r(T_j)$	$\eta(T_j)$
1	CART	0,032	87	86	6
2	C4.5	0,027	95	94	7
3	Эволюционный метод	0,014	61	57	7

Как видно, предложенный эволюционный метод обеспечил построение дерева решений с лучшими аппроксимационными свойствами, при этом обладающего меньшим количеством узлов.

Заключение

В работе решена задача построения диагностических моделей на основе деревьев решений.

Научная новизна работы заключается в том, что:

– предложен метод редукции обучающей выборки, который в отличие от существующих не только исключает малоинформативные признаки, но и удаляет из выборки экземпляры, не существенные для построения диагностических моделей, что позволяет синтезировать более простые модели, обеспечивающие приемлемые аппроксимационные и обобщающие свойства, затрачивая при этом меньше времени на процесс их идентификации;

– разработан метод синтеза деревьев решений на основе эволюционного поиска, который основан на стохастическом подходе и не использует жадную стратегию поиска, позволяет строить деревья решений с приемлемыми обобщающими и аппроксимационными свойствами, содержащие при этом незначительное количество узлов, что в свою очередь повышает интерпретабельность синтезированных моделей.

Практическая ценность результатов работы состоит в том, что решена задача синтеза модели диагностической модели коэффициента упрочнения деталей авиадвигателей.

Литература

1. Субботін С.О. Неітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей : [монографія] / С.О. Субботін, А.О. Олійник, О.О. Олійник ; [під заг. ред. С.О. Субботіна]. – Запоріжжя : ЗНТУ, 2009. – 375 с.
2. Rokach L. Data Mining with Decision Trees. Theory and Applications / L. Rokach, O. Maimon. – London : World Scientific Publishing Co, 2008. – 264 p.
3. Quinlan J.R. Induction of decision trees / J.R. Quinlan // Machine Learning. – 1986. – № 1. – P. 81-106.
4. Classification and regression trees / L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone. – California : Wadsworth & Brooks, 1984. – 368 p.
5. Quinlan J.R. Decision trees and decision making / J.R. Quinlan // IEEE Transactions on Systems, Man and Cybernetics. – 1990. – № 2 (20). – P. 339-346.
6. Прогрессивные технологии моделирования, оптимизации и интеллектуальной автоматизации этапов жизненного цикла авиадвигателей : [монография] / [Богуслаев А.В., Олейник Ал.А., Олейник Ан.А. и др.]. – Запорожье : ОАО «Мотор Сич», 2009. – 468 с.
7. Олійник А. О. Побудова нейромоделей для прийняття діагностичних рішень на основі еволюційної оптимізації / А.О. Олійник, С.О. Субботін // Автоматика-2009 : Шістнадцята Міжнародна науково-технічна конференція, 22 – 25 вересня 2009 р. : тези доповідей. – Чернівці, 2009. – С. 377-378.
8. Олейник А.А. Эволюционный метод структурно-параметрического синтеза нейронных сетей / А.А. Олейник, С.А. Субботин // Адаптивні системи автоматичного управління. – 2008. – № 12 (32). – С. 96-103.

Literatura

1. Subbotin S.O. Zaporizhzhya: ZNTU. 2009. 375 p.
2. Rokach L. London: World Scientific Publishing Co. 2008. 264 p.
3. Quinlan J.R. Machine Learning. 1986. № 1. P. 81-106.
4. Breiman L. California: Wadsworth & Brooks. 1984. 368 p.

5. Quinlan J.R. International Journal of Man-Machine Studies. 1987. № 27 (221). P. 221-234.
6. Quinlan J.R. Transactions on Systems, Man and Cybernetics. 1990. № 2 (20).P. 339-346.
7. Boguslaev A.V. Zaporizhzhya: Motor Sich. 2009. 468 p.
8. Oliinyk A. Automatics-2009: the sixteenth international scientific and technical conference, September 22-25, 2009. Chernovtsy. 2009. 377-378 p.
9. Oliinyk A. O. Adaptive systems of automatic control. 2008. № 12 (32). P. 96-103 .

Є.О. Гофман, А.О. Олійник, С.О. Субботін

Еволюційний метод синтезу дерев рішень

Досліджено еволюційний підхід до побудови дерев рішень. Розроблено метод редукції навчальної вибірки, що дозволяє виключати малозначимі та малоінформативні ознаки. Запропоновано еволюційний метод побудови дерев рішень. Проведено експерименти з синтезу діагностичних моделей на основі запропонованих методів.

E.A. Gofman, A.A .Oliinyk, S.A. Subbotin

Evolutionary Method of Decision Trees Synthesis

Evolutionary approach to constructing decision trees is investigated. A method for sample reduction, allowing to exclude irrelevant and uninformative features is created. Evolutionary method for constructing decision trees is proposed. Experiments on the synthesis of diagnostic models based on the proposed methods are conducted.

Статья поступила в редакцию 01.04.2011.